# Simple combinatorial considerations challenge Ruhlen's mother tongue theory

Louis-Jean Boë[1], Pierre Bessière[2], Nadia Ladjili[1], Nicolas Audibert[1]

**[1] Institut de la Communication Parlée (ICP), CNRS UMR 5009, Université Stendhal, INPG**
**[2] Informatique Graphique, Vision, Robotique (GRAVIR), CNRS UMR 5527**
**Grenoble, France**

## INTRODUCTION

In their proposal for the origin of internal structure of word forms, Peter MacNeilage and Barbara Davis (2000) asked whether patterns they observed in infants were relevant for historical linguistics. They tested this hypothesis on the list of the 27 Proto-World roots or Global Etymologies (henceforth GEs), proposed in 1994 by Merritt Ruhlen and John D. Bengtson (henceforth R&B). These GEs are presumed to derive from a common source, an ancestor language spoken 50,000 years ago. MacNeilage and Davis observe:

> This study shows that a corpus of proto-word forms shares four sequential sound patterns with words of modern languages and the first words of infants. Three of the patterns involve intrasyllabic consonant-vowel (CV) co-occurrence: labial (lip) consonants with central vowels, coronal (tongue front) consonants with front vowels, and dorsal (tongue back) consonants with back vowels. The fourth pattern is an intersyllabic preference for initiating words with a labial consonant-vowel-coronal consonant sequence (LC) … Remarkably, all three CV co-occurrence patterns favored by infants and languages are strongly favored, even in this extremely small protolanguage corpus [of GEs]. And the LC sequence is much more frequent than the CL sequence as well. (MacNeilage, Davis, 2000:527, 529).

And MacNeilage and Davis concluded from these observations:

> If the finding of not only the three CV co-occurrence patterns but also the LC effect in infant, language, and proto-language corpora means that these patterns are indeed basic to the origin of speech, then the controversial method of "multilateral comparison," pioneered by Greenberg (1963, 1987) and used by Bengtson and Ruhlen to construct their proto-language corpus, gains validity. (MacNeilage and Davis, 2000:529-530).

In this paper, we (1) Review certain key points in the debate and the issues at stake with respect to the existence of a single mother tongue for all the world's spoken languages.[1] (2) Review the methodology

[1] We note that the world's sign languages do not share a common origin with the world's spoken languages. The well-documented history of the emergence of Nicaraguan Sign Language in the 1970s makes this clear.

and the linguistics choices, both lexical and phonological, made by Ruhlen and Bengtson in order to arrive at the 27 Global Etymologies. (3) Propose a combinatorial evaluation of the lexical similarities observed by R&B between the lexicons of different languages distributed among different families. Are these similarities or are they not due to chance? In other words, can we reject the null hypothesis? (4) Demonstrate that the similarities observed between the lexicons of the different languages of the world simply reflect constraints on speech production and perception that are common to the human species and do not prove the existence of an original language.

## ORIGIN OF LANGUAGES AND ORIGIN OF MAN

For centuries, theories addressing the origin of Man and the origin of Languages have been closely linked. In the 17th century, an earliest interest in the origin of language and the identity of mankind was rekindled and became a dominant subject of discussion, but in the orthodox view, the Bible was still the main source of information about the earliest history of the earth and about mankind. It was believed that the earth, mankind and human language along with mankind, were no older than about six thousand years. By the turn of the 19th century, with the simultaneous developments of comparative philology and anthropology, the question of the origin of man and language had been revisited with new perspectives. In Germany (see E. Haeckel and A. Schleicher) and in France (see P. Broca and A. de Quatrefages) the relationship between linguistics and anthropology had always been very close. More generally, a circular argumentation was already appearing. On the one hand, linguists used anthropological hypotheses to validate their assumption of monogenesis for language. On the other hand, anthropologists referred to the hypothesis of the mother tongue to corroborate the assumption of monogenesis for mankind. Thus, in the

---

Kegl, J., Iwata, G. (1989). Lenguaje de Signos Nicaragüense: A pidgin sheds light on the "creole?" ASL. Proceedings of the Fourth Annual Meeting of the Pacific Linguistics Conference. University of Oregon, Eugene.

Senghas, A. (1995) The Development of Nicaraguan Sign Language via the Language Acquisition Process. D. MacLaughlin & S. McEwen (eds.), Proceedings of the Boston University Conference on Language Development 19, 543-552. Boston: Cascadilla Press.

19th century, de Quatrefages (1888), a “fixist” anthropologist took the theory of linguistic monogenesis and used to support that of the uniqueness of the human species.

And a century later, Ruhlen (1994) considered that his claim of a unique origin of languages was corroborated by certain work in population genetics (Cavalli Sforza et al., 1988 ; Cavalli Sforza, 1991). We believe that the debate will be made clearer if the questions of the origin of Man and of the origin of languages are dissociated. Even if we hypothesize the unique origin of modern Man, it is entirely possible that language appeared later in different populations. As Broca (1861-1862) emphasized, more than a century ago, the central question joining anthropology and linguistics is not that of the origin of languages, but that of the language faculty and its cerebral localisation. If there is a unique point, it is that the language faculty is specific to Man; whether there is a unique origin for languages is separate problem.

In our current state of knowledge, both the question of the origins of Man and the question of the origin of languages seem to be “ill-posed” problems in the mathematical sense of the term (as proposed by Hadamard in 1923). A problem is well-posed if a solution exists, the solution is unique, and the solution depends continuously on the initial data (stability property). A problem is ill-posed if it fails to satisfy at least one of the conditions. However, theories on the origin of Man are not stable; they rely too much on the discovery of new fossils. Each new discovery casts doubt on them, as recently occurred with the discoveries of the Abel fossil (Chad, 1995), Toumaï (Chad, 2001), and the Man of Flores (Indonesia, 2003). With the origin of language, the problem is complicated by the fact that there is no fossilized word, no recording, and no trace of writing dating to tens of thousands, much less hundreds of thousands, of years ago. And we know very well that the phonetic form of a modern word cannot reliably tell us about its phonetic form centuries ago. From a chronological point of view, 5,000 years is a possible leap back into the past for Indo-European constructions, but it is impossible to go back 50,000 years to reunite the languages of the peoples then in Australia, and it is impossible to go back 500,000 years to classify populations by DNA.

## MULTILATERAL OR MASS COMPARAISON

In the traditional comparative approach of historical linguistics, a relationship between two languages is revealed by finding words with different regular diachronic phonological correspondences and demonstrable affiliations in meaning. This principle of the *regularity of sound correspondences* it is basic to the sciences of etymology and comparative linguistics.

Greenberg's work uses a radically different approach. Greenberg's first major work was the genetic typology of the languages of Africa, published in the *Southwestern Journal of Anthropology* in 1949-50. The first principle (Greenberg 1949:79-83), is the exclusion of typological features (phonological, grammatical or semantic patterns) from genetic classification. Instead, "the arbitrary pairings of form and meaning, in both morphology and lexicon, provide the best evidence for genetic classification." The second principle is the exclusion of nonlinguistic evidence from the establishment of linguistic genetic families (Greenberg 1950a:57, 58).

After African language classification, Greenberg applied his methodology to the study of the languages of the Americas, and in Australia, he identified one widespread family, which he called *General Australian* and a large number of small families. These classifications were severely criticized and in response, Greenberg formulated a third principle, the simultaneous comparison of the full range of languages or *mass comparison*, later *multilateral comparison* (Greenberg 1954a:406-8). Contrary to the traditional comparative approach, the *multilateral approach* consists in looking at many languages across a few words rather than at a few languages across many words. This technique excludes historical linguistic history.

In fact, this method is neither original, nor recent, and it is not a revolution. In the 19th century, Antonio Balbi used linguistics in compiling his *Ethnographic atlas of the globe or classification of the modern and ancient peoples according to their languages*. For his classifications, he used 26 roots: *sun, moon, day, earth, water, fire, father, mother, eye, head, nose, mouth, tongue, tooth, hand, foot*, and the numbers from *one* to *ten*. As a by-product of his classifications, he uncovered similarities among words that did not seem to him to be due to chance, but which nevertheless did not allow him to consider that the words belonged to a single language family.

In this parallel between geographic boundaries and ethnographic [linguistic] boundaries, we have always abstracted from these analogies of form and of roots that one encounters more or less in almost all the languages of the globe, analogies which are too numerous to be reasonably attributed to chance, whereas they are not numerous enough, in other respects, to allow the ethnographer to classify all the languages he discovers in the same family. (Balbi, 1826) (our translation).

In 1952, before Greenberg had proposed his multilateral comparison method, Swadesh had developed his glottochronology using word lists as a tool for identifying linguistic 'stocks' worldwide. Using a single criterion of relatedness, that is the percentage of shared cognates, Swadesh proposed a classification of the languages of the Americas (1959) and also of the languages of the world (1962, 1971). Greenberg's classificatory work is in the same vein as that of Swadesh: he applies a single criterion of relatedness, a much looser version of Swadesh's criteria (measure of resemblance in core vocabulary), to a whole hemisphere.

## "MEGALOCOMPARISONS"

Following a suggestion from Dyen (1987) to provide evidence for the mother tongue, Ruhlen accepts that Greenberg's method of mass comparison of all the world's languages at once is valid without limitation:

Although the reduction of the world's roughly 5,000 languages to a mere 17 families would be considered a radical step by a good number of contemporary linguists, there are indications that even further consolidation is not only possible, but necessary. (…) In seeking a more comprehensive classification, one must avoid the trap of binary comparison, which at level of research seems so tempting. Rather, one should adopt a global approach, taking into account *all* of the world's languages, as Dyen (1959:546) suggested. (Ruhlen, 1987:258).

The attempt of Ruhlen can be considered to be a megalocomparison.

MICROCOMPARISON can be practiced on close-knit families like Romance, … a time-depth of no more than 2000 years. MACROCOMPARISON is appropriate for farflung by demonstrably valid groupings like Indo-European … , with time-depth of up to about 6000 years. MEGALOCOMPARISON takes on any more remote relationship, where sound-correspondance are not regular and putative cognates are few, so that chance rivals genetic relationship as the explanation for perceived similarities. (Matisoff, 1990:108).

### *The Global Etymologies of Ruhlen and Bengtson*

R&B's thesis (Bengtson and Ruhlen 1994a, 1994b, Ruhlen 1987, 1994a, 1994b, 1997) states that all the world's spoken languages come from a universal language spoken by a pre-proto sapiens 50,000 years

ago. This theory is backed up by a methodology which enables them to look for and find phonological and semantic equivalencies between words of different languages. In the end, these equivalencies enable him to make comparisons from a set of 32 families and 1316 languages and proto-languages (on average 41 languages per family). Finally they propose 27 GEs and, for each of them, the most general meaning and the phonological shape (Table 2). GEs have been presented as evidence for "proto-world". The results are listed in a database (Ruhlen, 1994a, 1997).

| | | | |
|---|---|---|---|
| **1**. ***mother older female*** AJA | **2**. ***knee to bend*** BU(N)KA | **3.** ***ashes dust*** BUR | **4.** ***nose To smell*** ćUN(G)A |
| **5.** ***hold (in the hand)*** KAMA | **6.** ***arm*** KANO | **7**. ***bone*** KATI | **8.** ***hole*** K'OLO |
| **9.** ***dog*** KUAN | **10.** ***who?*** KU(N) | **11.** ***woman*** KUNA | **12.** ***child*** MAKO |
| **13.** ***to suck Nurse, Breast***; MALIQ'A | **14.** ***to stay (in a place)*** MANA | **15.** ***man*** MANO | **16.** ***to think (about)*** MENA |
| **17.** ***what?*** MI(N) | **18.** ***two*** PAL | **19.** ***to fly*** PAR | **20.** ***arm*** POKO |
| **21.** ***vulva*** PUTI | **22.** ***leg Foot*** TEKU | **23.** ***finger one*** TIK | **24.** ***earth*** TIKA |
| **25.** ***leg Foot*** TSAKU | **26.** ***hair*** TSUMA | **27.** ***water*** ʔAQ'WA | |

***Table 2.*** *Global Etymology with their most general meanings and phonological shapes. We note that almost all of the GEs are taken from Swadesh lists (see Appendix 1). The GEs 6 and 20 on the one hand, and 22 and 25 on the other, differ only in their phonological form.*

Although they do not note this fact, R&B use a subset of words from Swadesh lists in the Global Etymologies they propose (21 GEs out of 27 are from Swadesh lists; see Appendix 1). These lists are available for more than a thousand languages and we have compiled a database for 361 languages from five families (Ladjili, 2005) using data from two websites.[2]

[2] For 95 Indo-European languages: Dyen et al., 1992, http://www.ntu.edu.au/education/langs/ielex/IE-DATA1

Over 1200 languages: "The Rosetta Project is a global collaboration of language specialists and native speakers working to develop a contemporary version of the historic Rosetta Stone," http://www.rosettaproject.org

## RUHLEN AND BENGTSON'S CRITERIA

To provide evidence for their Global Etymologies, R&B use a set of semantic and phonological correspondences.

### *Semantic equivalence*

Eco (1993) coined the expression "la furia dell'etimologia" to dub this frantic etymological hunting. The search for phonological similarities led R&B to extend the meaning of each root: they used, with an average or 24 semantic correspondences per Global Etymology. Thus, the first GE ***mother, older female*** is associated with the equivalent etymologies: *aunt, bitch = female dog, male dog, daughter, father, father's sister, female, girl, grandmother, husband's younger sister, father's mother, woman, woman's sister, wife*. The GE ***hole*** (number 8), is associated with: *anus, armpit, back, back part, breast, buttock(s), cave, crack, elbow= arm hole, fissure, grave, hiding place, hip, hole in the tree, hollow, incision, loins, nostrils, posterior, quiver, ravine, rear of army, river, shoulder, scoop out, to tickle, tickle a tired pig to make it go, tickling, valley*. And the GE ***to think (about)*** (number 16), is associated with: *abuse, admonish, appreciate, assume, be hungry, bewitch, brains, I can, command, conjecture, conjure, count, curse, desire, do you love me?, dream, folk song, guess, have friendly feelings, know, like, love, melody, memory, mind, name, order, possessed (applied to somnambulism), prayer, prefer, remember, remind, request, say, scold, seek, speak, story, summon, tale, talk, test, think, thought, try, understand, warn, wish, word*.

We do not intend to challenge the validity of these semantic equivalences (for critiques, see Bender, 1993 ; Salmon, 1992), but it is important to note that increasing the number of semantic equivalences increases the chances of "discovering" GEs. Nevertheless, we would expect Ruhlen to have selected world roots without any semantic overlap. It is therefore surprising to note that the list contains two pairs of identical roots (numbers 6 and 20, and 22 and 25), with the same general meanings ***arm*** and ***leg, foot***, different phonological shapes **KANO-POKO** and **TEKU-TSAKU,** and sharing five same meanings: *ankle, hip, hoof, thigh, upper leg*. We have questioned Ruhlen about this and in his response, he wrote:

One of the deciding factors (though not necessarily correct) was that there seemed to be a consistent difference between a stop and an affricate, though clearly the affricate could easily derive from the stop. Another factor was that in some families both roots seem to exist. For example, in Afro-Asiatic we find both *tak- 'walk' and *sAkA 'leg', and both roots seem to appear in Proto-Caucasian. In no way do I exclude the possibility that these two roots are really just one. (Ruhlen, personal communication, 2003).

The root numbers 10 **KU(N) who?** and 17 **MI(N) what?** share eleven meanings: do what, how many, what, what kind, what sort, when, where, who, who(ever), why.

More generally, Figure 1 presents semantic connections between the 27 GEs. Each link indicates that the two roots have at least one meaning in common. According to our calculations, 118 meanings are shared by two or more roots, and on average, a root shares a meaning with three other roots. If we put together the roots which share one or more meanings we obtain four semantic groups: (1) GEs 1, 9, 11, 12, 15; (2) GE 14; (3) GEs 18; (4) GEs 2-8, 10, 13, 16, 17, 19-27. The first group refers to people: *mother, woman (and bitch), man, child*. The second and third correspond to an action (*to stay*) and a number (*two*), and the last is a hodgepodge due to the "etymological fit of rage."

### *Phonological choices and equivalences*

To detect phonological similarities, Ruhlen uses well-known rules proposed by Indo-Europeanists of the 19th century, without taking chronology into account. All of the GEs (with the exception of GE 13) present a CVCV word structure. Ruhlen does not use complex rules to authenticate similarities between phonological shapes. Concerning consonant similarities he writes: "You don't need a Ph.D. in linguistics ... just common sense" (1994a: 18). He proposes main classes of similarities (1994a, p. 40) and we have systematically listed the equivalences used by R&B in their database (*On the Origin of Languages*, 1994, 277-330; *L'origine des langues*, 1997, 238-271). For consonants, they use ten macroclasses, with the following equivalences. (We will see later on that for R&B all vowels are equivalent.)

**1. /P B/** *p p$^{h}$ pʃ ph pj pjw pp pk pr pʔ*
*b bb bh br bw*
*ɓ ß cb ć ʃ f f$^{h}$ fl fr h hmb k kp lb m mb mp rf rp t tf th v w ʔb ʔl ʔmb ʔp ʔph*
For example, for **POKO** *boko faxa va ha spõka mbake*
**PAL** *hmbar pjeel wírre*

**PAR** *pʔ floˊ fri har*
**PUTI** *ʃapsi*

**2. /M/** *m mʳ mb md mg mj mk ml mlq mlqʷ mm mn mnj mnw mp mst mt mx*
*b bm ʃm g hm j jm ɬm ŋ ŋk nkʳ s sm t w xm zm ʔm ʔmq nʔ*

**3. /T D/** *t tʳ tʰ tʲ tˠ tʲtd tg th tk tl tᵬ tᵬʳ tᵬkh tn tr ts tt tx tz*
*d ð dʲ dh dʰ dr dᵬ dz*
*c c̃ ʃ ʃʃ ʃt ǧ gts h hs jt k kl ks ɬ lkʳ nd nǰ nt r ř rr rt s sʳ sʳk ss st stᵬ xts z ʔ ʔt*

**4. /TS/** *tˢ tˢ’ tˢk t ts tsh dtˢ’ dᶻ*
*ʃ ʃ’ ʃʰ ʃm ʃq’ g h j js ks kts ntˢ ŋs s stˢ zʔ ʔz*

**5. /N/** *n ñ nʲ ñj ñn ñt nb nc nʃ nʃ nd ng ngg ngk nh nʲ nj njg nkʷ nkx nl nm nn ns nt ntʳ nts ntj nw nz nʔx*
*b c ʃ ʃt ʃʳ d dn g h hh hm j jd jjn jn k kʳ kʷ kn kt l m mn ŋ ŋg ŋk p q r rn s sk t tʲ ts tsw tt w x z ʔ ʔn ʔt*

**6. /K G (N)K (N)G/** *k k’ kb kʃ kʰ kh khs kj kkh kk kkw kl kl’ km kⁿ kn kp kr ks kt kth kts kv kʷ kw k’w kx nk ŋk nk’ ŋkw*
*g gb gg gh gl gn gr gt gʷ gw gwʔ ng ngg ngk nǧ njg ŋ ŋg*
*b bs c cl ʃ ʃb ʃk ʃp ʃml ff d ɤ ɤ’ ɤŋ ɤw h hk hw jj jk kʔ l lk lp m mlj n nc nd nj nn ns nt nw p pp pk q qq q’ qk qn qs qˀʷ qʷ qw r rs s sh skw sp sq t tʰ tk tˡ tˢk tˢkw ts v w wǧ wɤ wh wkʰ x xb xj xk xl xn xp xs xt xtˢ xʷ z ž zk ʔg ʔk ʔp ʔtks ʔts ʔw !k*

**7. /Q’ Q’W/** *q qʳ qʷ qʎ qv*
*ʃ ʃʃ ʃʳ ʃʲ ʃt d g ǧ gʷ h hqʰ hw hk k kk kʳ kʳkʳ kʰ kh kʲs kʳʳw kk kl ks kʷ kw ʎqʳ rk ʁʷ st w x xk xvb xʷ xʷʔ žt ʔ ʔkʔk*

**8. /L/** l l’ lh lʲ lk ll lqm lst
*gl j kr n ng r rd rn rr rs*

**9. /R/** *r rb rc rd rf rɤ rɤw rk rp rr rs rt rx*
*b bʔd d dr l n nn sk wr*

**10. /J/** *j*
*h k kw l n ng r ts*

We note that R&B do not hesitate to use equivalences even among these macroclasses. We have listed these equivalences and present them in Figure 1.

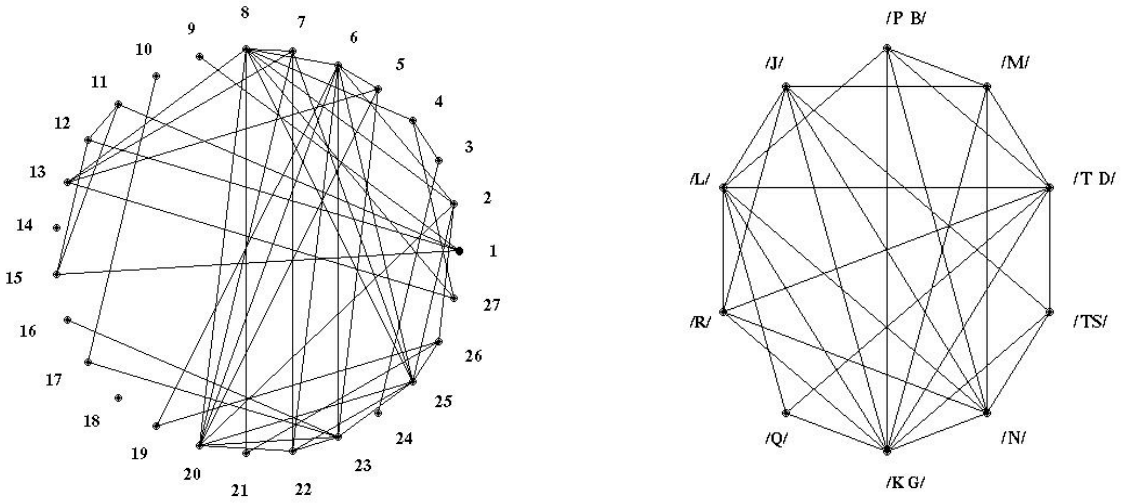


***Figure 1.*** *Right: Connections between roots sharing at least one meaning (an average of three). Left: Similarities between phonemes sharing at least one confusion (for example, /P B/ can be confused with /M/ or with /T D/ or with /N/, even though they are different phonemes).*

For vowels, it is well known, for example in Indo-European, that vowel quality can change over the course of time, by the processes of *umlaut*, metaphony or inflection. It is therefore not surprising that R&B do not take vowels into account and rather consider them all to be similar. Thus, for the /A/ of the GE **PAL**, R&B propose in addition to /a/, the vowels i e ɛ ə o u / as in *firi, pilia, wi, prin, peele, peia, ferfir, pɛle, wər, pol, poya, von, pula, brue, huave, bu, bur*.

For the GE **TIK**, R&B considers the following phonological shapes as equivalent: *atgu, ʃiʃ, ɗeʕ, digitu, dlianŋ, ɗɔkku, itygin, iʃaki, motook, otejji, řak, sik, taihwo, təgu, tku, tsʷiʕʷ, sakwe, zɛkatikkuagpaa*. The vowel /I/ of this GE is judged to be equivalent to nine other vowel phonemes as indicated in Table 3.

| **i** | **e** | **ɛ** | **a** | **y** |
|---|---|---|---|---|
| **38%** | **20%** | **1%** | **15%** | **1%** |
| **ɨ** | **ə** | **u** | **o** | **ɔ** |
| **1%** | **1%** | **8%** | **13%** | **1%** |

***Table 3****. Vowel phonemes and percentages of occurrences. corresponding to phoneme /I/ of the Global Etymology* ***TIK.***

**Such strong phonological similarities for consonants and vowels lead to only 242 different possible (C)VC(V) phonological shapes, if we allow the presence or absence of the first consonant and of the final vowel (11 × 11 × 2).**

### *Criteria of emergence of a GE*

To support the existence of 27 GEs, R&B (Ruhlen, 1994 :277-328 ; Ruhlen, 1997 : 233-271) indicate that:

**the 27 world roots ... are represented in at least six of these families, but a root is represented on average in 12 families, and the most widespread one, KU(N), 'who?', is represented in 23 or 24 families (Ruhlen, 1997: 234) (our translation).**

Figure 2 presents a schematization of this criteria: to belong to same GE, at least six equivalent phonological forms out of 24 synonyms must be present in at least six languages from six different families.

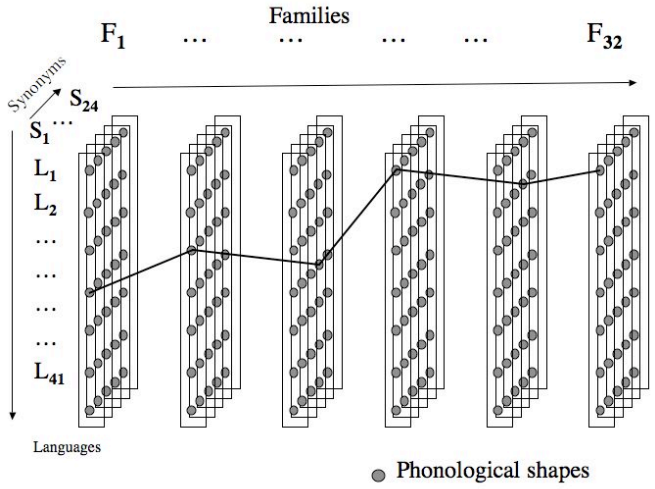


***Figure 2.*** *Ruhlen and Bengtson's criteria: for each Global Etymology find the same phonological form + meaning, including 24 synonyms (S), in six languages (L) belonging to 6 different families (F).*

Is this conclusive proof of the existence of a mother tongue or could it be the result of mere chance or of lexical constraints?

## TESTING THE NULL HYPOTHESIS

There is currently a considerable literature detailing how criteria for determining both a semantic and a phonological match are almost entirely lacking, and there are numerous publications arguing against multilateral comparisons (Campbell, 1988, 1998a 1998b, 1999; Guy, 1995; Matisoff, 1990; McMahon and McMahon, 1995; Nichols, 1996; Rankin 1992, Ringe, 1992, 1995, 1996, 1998, 1999; Trask, 1996). Aitchison provides an excellent summary of the problems, which we discuss further below:

> Chance resemblances are easy to find among different languages if only vague likenesses among shortish words are selected… sounds change radically over the centuries. Words which existed so long ago are unlikely to have survived in anything like their original state… Taboo is a further problem… the 'lucky dip' approach does not make any attempt to eliminate accidental correspondences, nor does it control for phonetic probability or tabo… meanings tend to be reduced to fairly simple, straightforward items, with a limited number of phonetic shapes. In these circumstances, chance similarities are likely to play a worryingly high role, and this 'mass comparison' method is unlikely to stand the test of time. (Aitchison , 1996: 173)

In more damning critiques, many linguists even consider that the Global Etymologies "cannot teach us anything about the origins of human languages" (Bender, 1993; Hock, 1993; McWhorter, 2001; Picard, 1998; Rosenfelder, 1999; Salmon1992a, 1992b, 1997).

In fact any demonstration of a relationship between languages depends largely on finding words of similar phonological shape and roughly equivalent meaning in the languages considered. However it must be shown that the similarities observed could not have arisen by chance. Unfortunately Ruhlen does not take this precaution. We must therefore determine whether the observed similarities give us reason to reject the null hypothesis, that is, the hypothesis that the similarities are merely a product of chance factors. As Hurford puts it:

> As linguists like Larry Trask, Don Ringe and Lyle Campbell, to name but a few, loudly insist, no good answer has yet been given to the charge that the correspondences noted by the long-range reconstructionists are not above the chance level. In other words, no effort has been put into rejecting the null hypothesis. (Hurford, 2003).

### *Lexical homogeneity/heterogeneity of families*

The number of different phonetic forms for the same root (*Ndiff*) in a given family seems to us to be an essential lexical parameter. This number gives us an idea of the cohesiveness and even the existence of families. The value of *Ndiff* must of course be greater than 1; otherwise there would only be a single

language per family. But on the other hand, *Ndiff* cannot be as great as the number of languages in the same family; otherwise, the very existence of the family would be called into question. *Ndiff* is therefore an important piece of information about linguistic typology that has led the experts to classify languages by family: a compromise between the homogeneity of the family and the existence of different languages for the same family. The value of *Ndiff* can only be evaluated using counts realized with databases. We have performed database counts using Swadesh lists for the following five families: Afro-Asiatic, Andamanese, East Papuan, Austronesian and Indo-European, and have found a different number of languages per family. Figure 3 presents the average value for the GEs proposed by R&B.

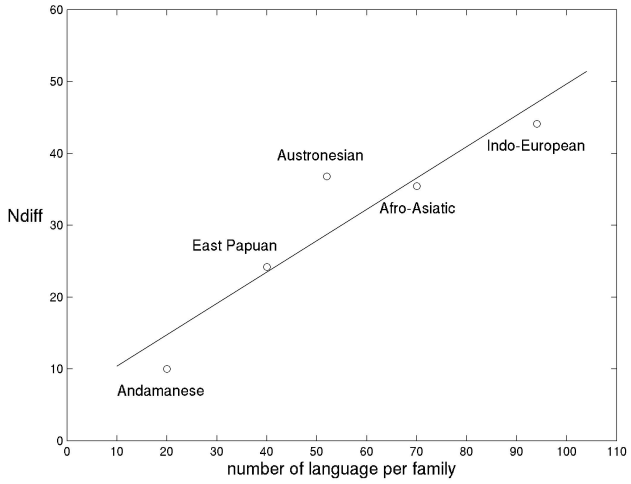


***Figure 3.*** *Mean values for the number of different phonological shapes Ndiff for Ruhlen's 19 GEs in five families in relation to the number of languages per family that we analyzed.*

The mean values for *Ndiff* by family vary between 9 for Andamanese and 45 for Indo-European, and we observe a significant correlation ($R = 0.92$ ; $p = 0.022$) with the number of languages per family. In choosing a great number of languages per family, we thus proportionally increase the chance of finding similarities; Ruhlen worked with 41 languages per family, which corresponds to an *Ndiff* of approximately 15, and this number increases with the number of synonyms. To illustrate this, we take the case of the GE ***who*****,** for which Ruhlen and Bengtson use synonyms like ***what, when, where, how,*** and ***if***. These synonyms are also found in the Swadesh lists. For the five families, we have thus calculated the *Ndiff*

values for the GE ***who*** after the addition of 1 to 5 of these synonyms. We note that *Ndiff* increases and tends towards a different asymptote for each family (Figure 4): between 45 and 180. With 24 synonyms and 41 languages per family, R&B have an *Ndiff* of at least 60.

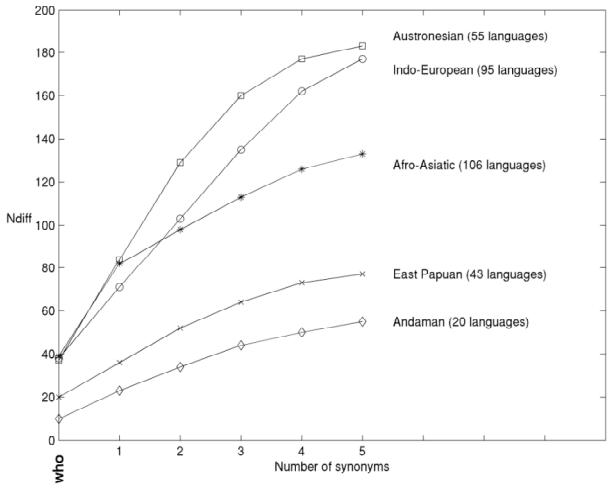


***Figure 4***. *Values of* Ndiff *for the GE* ***who*** *after the addition of synonyms (results for five families). Ndiff increases with the number of semantic correspondences and depends on the number of languages per family that we analyzed for this GE : Afro-Asiatic is the most homogeneous family, and Austronesian the least homogeneous.*

The GEs ***water, mother*** and ***dog*** have the lowest *Ndiff* values and ***man, woman***, and ***child*** the highest values. These tendencies should be confirmed with a greater number of families. But we are not surprised to find ***mother*** among the GEs with low values. It has been well-known for centuries that nursery forms (like mama, nana, papa, dada, kaka) are similar across the world's languages. Murdock (1959) listed 531 terms for ***mother*** in different languages and concluded that the data :

[c]onfirm the hypothesis [of] a striking convergence in the structure of these parental kin terms throughout historically unrelated languages. (Murdock, 1959).

And Jakobson in his paper Why *« MAMA » and « PAPA » ?* proposed the following explanation:

Since the mother is, in Grégoire parlance, *la grande dispensatrice*, most of the infant's longings are addressed to her, gradually turn the nasal interjection into a parental term, and adapt its expressive make-up to their regular phonemic pattern. (Jakobson, 1962:543).

### *Definition, notation and calculation*

R&B find that for each GE, similar phonetic forms are found in at least six families. Is this a conclusive evidence of the existence of a mother tongue or could it be the result of mere chance? To answer this question we propose a very simple combinatorial computation:

Let *Nge* be the number of *GEs* taken into account in the reasoning.

Let *Nm* be the mean number of meanings per *GE*. In R&B's study, *Nm* = 24.

Let *Ndiff* be the number of different phonological shapes associated with a given *GE* in a given family. Of course, the bigger *Nm* is, the bigger *Ndiff* is. According to thorough our analysis, *Ndiff* is at least 10 (see Figure 4).

Let *Nph* be the total number of different possible phonological shapes. With Ruhlen's phonological similarity rules, for a CVCV structure, *Nph* = 182 (see § 3).

Let *Nf* be the total number of considered language families; In Ruhlen's demonstration *Nf* = 32.

Finally, let *f* be the minimum number of families that contain for a given *GE* at least one phonological shape in common with another family.

In the following calculations, we assume that there are neither semantic equivalences between the 27 roots nor phonological equivalences between the 10 macroclasses.

For our demonstration, let us construct a table in which each column corresponds to a language family and each row to a possible phonological shape. Such a table has *Nf* columns and *Nph* rows. For a given *GE*, let us enter 1 in each of the *Nf* x *Nph* cells when a phonological shape appears in a given family associated with this *GE*. If the phonological shape does not appear, let us enter 0.

A small instance of such a table is given by Table 1 of Appendix 2 for *Nf* = 10, *Nph* = 25 and *Ndiff* = 5 (all the columns contain five cells containing a 1).

We can permute the position of two columns without changing the association between the phonological forms and the families. The same is true if we permute two rows.

Thus we can group on the left all the columns corresponding to families that share a phonological shape with another family. Let us imagine that we have $\Phi$ such families. This leads to Table 2 of

Appendix 2 where $\Phi$ = 7. All the column on the right contain *n* cells containing a 1, and there are $Nf - \Phi$ such columns.

Afterwards, we can group at the top all the rows of these $\Phi$ families which contain at least one 1. Let us imagine that we have $\xi$ such rows. We obtain Table 3 of Appendix 2 where $\xi$ = 7. All the rows underneath should, because of how we have constructed the table, contain at most one cell with a 1. There are $Nph - \xi$ such rows.

We necessarily obtain the following inequality:

$$(Nf - \Phi) \times Ndiff \leq Nph - \xi$$

Obviously, $\xi \geq Ndiff$ so we obtain:

$$(Nf - \Phi) \times Ndiff \leq Nph - Ndiff$$

The formula is also true for the minimum possible value of $\Phi$ : *f*. This leads to:

$$(Nf - f) \times Ndiff \leq Nph - Ndiff$$

And we finally derive:

$$f = \frac{Nf - Ndiff}{Ndiff} \qquad (\text{if } f < 0 \text{ , } f = 0)$$

Figure 5 plots this minimum value of *f*. It reveals that for any given GE, as soon as *Ndiff* exceeds 10 there are at least 14 families that share similar phonologic shapes (with *Nph* = 200).

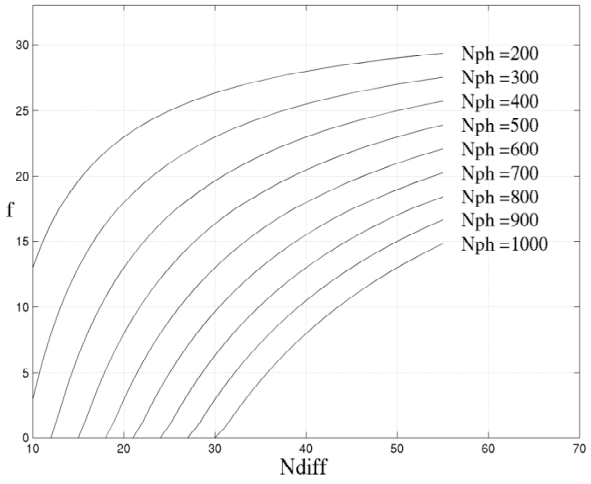

***Figure 5.*** *The minimum number of families f, that share similar phonologic forms, plotted as a function of Ndiff, the number of different phonological shapes for a given family and a given GE, and for different values of Nph.*

With 24 synonyms, or a minimum value of *Ndiff* = 60, R&B have a 100% chance of finding similarities among a couple dozen families. Even if they increase the number *Nph*, assuming a couple dozen consonant categories (*Nph* on the order of1000), they would find similarities in at least six families.

**The null hypothesis cannot be rejected: Ruhlen use too many equivalent meanings for each GE, and too many phonological equivalences for a too small number of different phonological shapes.**

In effect, by using the data base that we have compiled using Swadesh lists and by adopting Ruhlen's criteria of emergence (*find the same phonological shape + meaning including 24 synonyms, in at least six languages belonging to 6 different families*), we were able to find for the GE ***mother*** two other phonological forms (number 1) different from AJA: **MAMA** for the Sino-Tibetan, Afro-Asiatic, Niger-Congo, Indo-European, Austronesian, Indo-Pacific (Adamanese and East Papuan) and Amerind families ; **ANA** for Indo-Hitite, Ouralian, Turk, Sino-Tibetan, Niger-Congo, Afro-Asiatic, Austronesian, and Indo-Pacific. Similarly for the GE ***who*** (number 10), we find for the Basque, Niger-Congo, Afro-Asiatic, Daic, Indo-European, Japanese-Ryukyuen, Na-Dene, Sino-Tibetan families the phonological form **TA**, which is different from the form **KU(N)** proposed by R&B. **We can therefore conclude that by applying the methods of R&B, we provide evidence for several mother languages.**

## WHY SO MANY PHONOLOGICAL SIMILARITIES?

During almost half a century, the dominant theories in the field of language science (structuralism and generativism) have imposed an "inescapable hypothesis." From Saussure to Chomsky, the study of speech (*parole*) was relegated to a position of secondary importance, subordinate to the study of the language system (*langue*). Only at the beginning of the 1970s did a new approach (Lijencrants, Lindblom, 1972; Stevens, 1972) arise which places speech at the origin of the structuring of the phonological systems of the world's languages — at least their great trends — and which suggests a first modeling in this domain. We

shall say that "substance," that is the speech production and perception processes, "informs" the phonological systems of the world's languages.

These processes are at the origin of the organization of these structures, other (biological, sociological) processes being, of course, likely to intervene as well. A series of studies of this "substance based" phonology, in which Davis and MacNeilage participate for ontogeny, has shown that the general tendencies of consonant and vowel systems can be explained and modeled by production and perception constraints (see Schwartz et al., 1977b; Boë et al., 2005). The typologies of the phonological systems of the world's languages (Troubetzkoy 1939, then Crothers 1978, and more recently, Vallée 1994 and Schwartz et al., 1997a) revealed that languages use relatively limited choices among all the possibilities determined by simple combinatory rules. The UPSID (Maddieson, 1986; Maddieson, Precoda, 1989) (UCLA Phonological Segment Inventory Database) (317 then 451 languages and now 566) gathers phonological systems of the world's languages, sampling more or less uniformly all language families. Among all the available consonants, phonological systems use only a very small combination, although the possible combinations offered by the many categories of manner and place of articulation are quite large. Three places of articulation are the most common: bilabial, coronal, and velar (Table 3).

| | **Bilabial** | **Coronal** | Postalveolar | Palatal | **Velar** | Uvular | Pharyngeal | Glottal |
|---|---|---|---|---|---|---|---|---|
| STOPS | **99 *** | **100 *** | 6 | 16 | **99 *** | 13. | 1 | 48 |
| NASALS | **95 *** | **96 *** | 10 | 31 | **53** | 0 | 0 | 0 |
| FRICATIVES | **58** | **85** | 43 | 8 | 29 | 11 | 4 | **62** |
| AFFRICATES | 0 | **85 *** | **49** | 4 | 1 | 1 | 0 | 0 |
| APPROXIMANTS | **79** | **78 *** | 3 | **85 *** | **75** | 0 | 0 | 0 |

***Table 3***. *Percentages for places of a articulation by manner of articulation for the 451 languages of UPSID (Boë et al., 2000). In bold, the three or four places of articulation for each manner of articulation. Coronal includes dental, alveodental, alveolar. The * marks the macroclasses adopted Ruhlen & Bengtson.*

R&B's choice of 10 macroclasses of consonants corresponds almost exactly to the most common categories (with the exception of /Q/). It is therefore not surprising that this choice favors the phonological similarities of the GEs. Nor is the fact that MacNeilage and Davis observe that three CV co-occurrence patterns favored by infants and languages are strongly favored in data of R&B surprising. These main

tendencies reveal that all the speakers of the planet have the same vocal tract, the same auditory system, and the same principles of control, but it is not a proof of the existence of a unique mother tongue.

## CONCLUSION

We have demonstrated, by simple combinatorial considerations, that the Global Etymologies proposed for a proto-sapiens language in *The Origin of Languages* can be explained by random chance: The null hypothesis cannot be rejected. With the methodology used, Ruhlen and Bengtson had a 100% chance of finding 27 Global Etymologies common to the 32 families and thereby of validating their mother tongue hypothesis. They used too few Global Etymologies, too many equivalent meanings, too many languages per family, and too many phonological equivalences for a too small number of different phonological shapes.

Our demonstration is not proof that the hypothesis of monogenesis for spoken languages can be rejected, simply that the procedure adopted by Ruhlen and Bengtson has no combinatorial validity. It seems that the method used to reconstruct proto-languages is one of the weakest points of this "New Synthesis" promoted by L. Cavalli-Sforza, C. Renfrew and M. Ruhlen which brings together genetic, archeological and linguistic data in the reconstruction of human evolution.

Is it really possible to prove the existence of a mother tongue spoken 50,000 years ago with a simple comparison of phonological forms of modern spoken words? The MacNeilage and Davis program, which seeks to find universal sensorimotor principles, seems to us more founded, more promising in terms of development and results for the origin of speech. Furthermore, this program has the advantage of being independent of the question of the existence of a unique mother tongue.

## Acknowledgements

This research was conceived and started in the *Representation and Diffusion of Scientific Ideas in Speech and Language Sciences* project (managed by Louis-Jean Boë and Christian Abry, ICP, since 1999) and funded by the *Maison des Sciences de l'Homme-Alpes*. It was a part of research in the *Congruence* project (managed by Pierre Darlu, INSERM, since 2001), being part of the *Origine de l'Homme du Langage et des Langues* project (managed by Jean-Marie Hombert) funded by the CNRS. Thanks a lot to

Christian Abry, Paul Boersma, Lyle Campbell, Rob McMahon, James Mattisoff, Manu Mazer, Boyd Michailovsky, Diverson Mzemba, Don Ringe, and Pauline Welby for their assistance helpful criticism, suggestions and documentation.

## References

Aitchison, J. (1996) *The seeds of speech: language origin and evolution*. Cambridge: Cambridge University Press.

Bender, M. L. (1993.) Are Global Etymologies Valid? *General Linguistics* 33,191-219.

Bengtson, J.D., Ruhlen, M. (1994) Another look at *TIK 'Finger, One'. *California Linguistics Newsletter*, 24(2):9-11.

Boë, L.J. (1997) Sciences phonétiques et relations forme/substance : 1. Un siècle de ruptures, négociations et réorganisations. 2. Du poids de la substance sur la forme aux réarticulations scientifiques. *Histoire Epistémologie de la Linguistique, HEL* XIX, 1, 5-41, 2, 5-25.

Boë, L.-J., Abry, C., Cathiard, M., Schwartz, J.-L., Vallée, N., Badin, P. (2005). Comment les exceptions des handicaps révèlent les universaux phonologiques : contraintes visuelles et auditives des systèmes consonantiques des langues du monde. *Faits de langues*, 25 (Special issue «L'exception entre les théories linguistiques et l'expérience » ed. by I. Vilkou-Poustovaïa ), 175-190.

Boë, L.J., Schwartz, J.L., Vallée, N. (1994) The prediction of vowel systems: Perceptual contrast and stability. In " Fundamentals of Speech Synthesis and Speech Recognition, 185-213. Ed. by E. Keller. Chichester : John Wiley.

Boë L.J., Vallée N., Badin P., Schwartz J.L., Abry C. (2000) Tendencies in phonological structures: the influence of substance on form. *Bulletin de la Communication Parlée*, 5, 35-55.

Boë, L.J., Vallée N., Schwartz J.L., Abry C. (2002) The nature of vowel structure *Acoustical Science and Technology*, 23,4, 221-228.

Broca P. (1861) Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauche du cerveau. *Bulletins de la Société d'Anthropologie de Paris*, 2, 235-238.

Broca, P. (1862) La linguistique et l'anthropologie. *Bulletins de la Société d'Anthropologie de Paris*, III, 264-319.

Campbell, L. (1988) Review of Language in the Americas, by Joseph Greenberg. *Language* 64, 591-615.

Campbell, L. (1998a) *Historical linguistics*. Edinburgh: Edinburgh University Press. (1999, American edition: Historical linguistics. Cambridge, MA: MIT Press.)

Campbell, L. (1998b) Nostratic: a personal assessment. Nostratic: sifting the evidence, ed. by Brian Joseph and Joe Salmons, 107-152. Amsterdam: John Benjamins.

Campbell, L. (1999) Nostratic and linguistic palaeontology in methodological perspective. Nostratic: Evaluating a Linguistic Macrofamily, ed. by Colin Renfrew and Daniel Nettle, 179-230. Cambridge: The McDonald Institute for Archaeological Research.

Cavalli-Sforza, L.L. Piazza, A. Menozzi, P. Mountain, J. (1988) Reconstruction of Human Evolution: Bringing together Genetic, Archeological and Linguistic Data," *Proc. Nat. Acad. Sciences*, 85, 6002–6006, 1988.

Crothers, J. (1978) Typology and Universals of Vowel Systems. In *Universals of Human Language*. J.H. Greenberg (Ed.), 93-152, Stanford Univ.Press.

Dyen, I. (1959) Review of *Essay in Linguistics*, by Joseph Greenberg. *Language* 37, 527-552.

Dyen, I. Kruskal, J.B, Black,P. (1992) An Indoeuropean Classification: A Lexicostatistical Experiment. *Transactions of the American Philosophical Society*, 82, 5. http://www.ntu.edu.au/education/langs/ielex/IE-DATA1

Eco, U. (1993) *La ricerca della lingua perfetta nella cultura europa*, Roma-Bari: Laterza.

Greenberg, J. H. (1949) Studies in African linguistic classification: I. Introduction, Niger-Congo family. *Southwestern Journal of Anthropology* 5.79-100.

Greenberg, J.H. (1950a) Studies in African linguistic classification: IV. Hamito-Semitic. *Southwestern J. Anthropology* 6, 47-63.

Greenberg, J.H. (1950b) The patterning of root morphemes in Semitic. *Word* 6.162-81.

Greenberg, J.H. (1953) Historical linguistics and unwritten languages. *Annthropology today*, ed. A. L. Kroeber, 265-86. Chicago: University of Chicago Press. (Reprinted in Greenberg 1971b, 11-48.)

Greenberg, J.H. (1954) Studies in African linguistic classification VIII. Further remarks on method; revisions and corrections. *Southwestern Journal of Anthropology* 10.405-15.

Greenberg, J.H. (1955) *Studies in African linguistic classification*. New Haven: Compass Press.

Greenberg, J.H. (1963) *The languages of Africa*. Bloomington: Indiana University Press.

Greenberg, J.H. (1987) *Language in the Americas*. Cambrige: Cambridge University Press.

Greenberg, J.H. (1988) Language in the Americas. *Language*, 64, 591–615.Jakobson, R. (1962) 'Why "mama" and "papa"?' In Jakobson, R. *Selected Writings, Vol. I: Phonological Studies*, pp. 538–545. The Hague: Mouton.

Grégoire, A. (1937) *L'apprentissage du langage*. Bibliothèque de la Faculté de Philosophie et de Lettres de l'Univ. de Liège, 73.

Guy, J. (1995) Merritt Ruhlen: On the Origin of Languages, book review. *Anthropos: revue internationale d'ethnologie et de linguistique*, 90, 638–639.

Hock, H.H. (1993) SWALLOW TALES: chance and the "world etymology" MALIQ'A 'swallow, throat'. *Chicago Linguistic Society* 29.215-9.

Hurford, J.R. (2003) Review of Michael C. Corballis, From Hand to Mouth: the origins of language, Princeton: Princeton University Press, 2002. *Journal of Linguistics*, 39, 201-216. Cambridge University Press. http://www.ling.ed.ac.uk/~jim/corballisrevu.html

Jakobson, R. (1929) Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slaves. Travaux du Cercle Linguistique de Prague, 2. [Reprinted in Selected writings of Roman Jakobson, volume 1: Phonological studies, 538-545. The Hague: Mouton, 1962, first edition.]

Ladjili, N. (2005*) Constitution d'une base de données de listes lexicales de Swadesh. Les différences phonétiques intrafamille*. Master Industrie de la Langue. Université Stendhal, Grenoble.

Liljencrants, J., Lindblom, B. (1972) Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language* 48, 839-862.

MacNeilage, P.F. (1998) The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, 499-511

MacNeilage, P.F. and Davis, B. L. (2000) On the origin of internal structure of word forms. *Science*, 288, 527-531.

Maddieson, I. (1986) *Patterns of sounds*. Cambridge: Cambridge University Press. 2nd ed.

Maddieson, I., Precoda, K. (1989) *Updating UPSID*. UCLA Working Papers in Phonetics 74, 104-111.

Matisoff, J.A. (1990) On megalo-comparison: a discussion note. *Language* 66, 106-20.

McMahon, A., McMahon, S. (1995) Linguistics, genetics and archaeology: internal and external evidence in the Amerind controversy. *Transactions of the Philological Society* 93.125-225.

McWhorter, J. (2001) The power of Babel: a natural history of language. New York: Time Books.

Murdock, G.P. (1957) Cross-language parallels in parental kin terms. *Anthropological Linguistics* 1 (9), 1-5, 1957.

Nichols, J. (1996) The comparative method as heuristic. The comparative method revised, ed. by Mark Durie and Malcolm Ross, 39-71. Oxford: Oxford University Press.

Picard, M. (1998) The Case against Global Etymologies: Evidence from Algonquian. *International Journal of American Linguistics* 64, 141-7. [Published also 1995, On the nature of the Algonquian evidence for global etymologies. *Mother Tongue* 24. 50-54.]

Rankin, R.L. (1992) Review of Language in the Americas, by J.H. Greenberg. *International Journal of American Linguistics* 58.324-51.

Renfrew, C. (1987) *Archaeology and language : The puzzle of indo-european origins*, London: Jonathan Cape, 1987.

Ringe, D.A. (1992) On calculating the factor of chance in language comparison. *Transactions of the American Philosophical Society*, 82, 1, 1–110, 1992.

Ringe, D.A. (1995) Nostratic" and the factor of change. *Diachronica* 12.55-74.

Ringe, D.A. (1996) The mathematics of 'Amerind'. *Diachronica* 13, 135–54.

Ringe, D.A. (1998) Probabilistic evidence for Indo-Uralic. Nostratic: sifting the evidence, ed. by Brian Joseph and Joe Salmons, 153-97. Amsterdam: John Benjamins.

Ringe, D.A. (1999) How hard is to match CVC-roots ? Transactions of the Philological Society 97, 213-244.

Rosenfelder, M. (1999) Deriving Proto-World with tools you probably have at home http://www.zompist.com/proto.html

Ruhlen, M. (1987) *A guide to the world's languages, volume 1: classification.* Stanford: Stanford University Press.

Ruhlen, M. (1994a) *The origin of language. Tracing the evolution of the mother tongue*, New York: John Wiley & Sons, Inc.

Ruhlen, M. (1994b) *On the origin of languages. Studies in linguistic taxonomy*, Stanford: Stanford University Press.

Ruhlen, M. (1997) *L'origine des langues*. Paris: Belin.

Salmons, J. (1992a) A look at the data for a global etymology: *tik 'finger'. Explanation in Historical Linguistics, ed. by Garry W. Davis and Gregory K. Iverson, 208-28. Amesterdam: John Benjamins.

Salmons, J. (1992b) Theory and practice of global etymology. *Proceedings of the 15th International Congress of Linguists* 1, 153-155.

Salmons, J. (1997) 'Global etymology' as pre-Copernican linguistics. *California Linguistic Notes* 25, 1, 5-7.

Schwartz, J.L., Boë, L.J.,Vallée, N., Abry, C. (1997a) Major Trends in Vowel System Inventories. *J. of Phonetics* 25, 3, 233-253.

Schwartz, J.L., L.J. Boë, L.J., N. Vallée, N., Abry, C. (1997b) The Dispersion-Focalization Theory of Vowel Systems. *J. of Phonetics* 25, 3, 255-286.

Stevens, K.N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data, In Human Communication: A unified view. 51-66. Ed. by E.E. Davis Jr. and P.B. Denes, Mc Graw-Hill, New-York .

Swadesh, M. (1952) Lexicostatistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96, 452-463.

Trask, R.L. (1996) *Historical Linguistics*, London: Arnold.

Troubetzkoy, N.S. (1939). Grundzüge der Phonologie. *Travaux du Cercle Linguistique de Prague* 7. [French translation by J. Cantineau *Principes de phonologie*. 1949, 1970, Paris : Klincksieck.]

Vallée, N. (1994) *Systèmes vocaliques : de la typologie aux prédictions*. Thèse de Doctorat en Sciences du Langage, Université Stendhal, Grenoble.

# Appendix 1

all
and
animal
**ASHES**
at
back
bad
bark
because
belly
big
bird
bite
black
blood
blow
**BONE**
breast
breathe
burn
**CHILD**
cloud
cold
come
correct
count
cut
day
die
dig
dirty
**DOG**
drink
dry
dull
dust
ear
**EARTH**
eat
egg
eye
fall
far
fat
father
feather
**HAIR**
few
fight
**FINGERNAIL**
fire
fish
five
float
flow
flower
**FLY**
fog
**FOOT**
forest
four
freeze
fruit
full
give
good
grass
green
guts
hand
he
head
hear
heart
heavy
here
hit
**HOLD**
horn
how
hunt
husband
I
ice
if
in
kill
**KNEE**
know
lake
laugh
leaf
left
**LEG**
lie
live
liver
long
louse
**MAN**
many
meat
moon
**MOTHER**
mountain
mouth
name
narrow
near
neck
new
night
**NOSE**
not
old
one
other
person
play
pull
push
rain
red
right
river
road
root
rope
rotten
round
rub
salt
sand
say
sing
scratch
sea
see
seed
sew
sharp
short
sit
skin
sky
sleep
small
smell
fear
smoke
smooth
snake
snow
some
spit
split
squeeze
stab
stand
star
stick
stone
straight
**SUCK**
sun
swell
swim
tail
that
there
they
thick
thin
**WOMAN**
**THINK**
this
thou
three
throw
tie
tongue
tooth
tree
turn
**TWO**
vomit
walk
warm
wash
**WATER**
we
wet
**WHAT**
when
where
white
**WHO**
wide
wife
wind
wing
wipe
with
worm
year
yellow

200-word Swadesh list. In bold upper case letters, the Global Etymologies of Ruhlen and Bengtson.

## Appendix 2

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 6 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| 7 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| 8 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| 11 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 17 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 18 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 20 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |

*Table 1*

| 1 | 10 | 3 | 8 | 5 | 6 | 7 | 4 | 9 | 2 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |

*Table 2*

| | 1 | 10 | 3 | 8 | 5 | 6 | 7 | 4 | 9 | 2 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 17 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 25 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| 10 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 14 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 6 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*Table 3*